CDC WONDER on the Web

Andrew Friede, MD, MPH, Patrick W. O'Carroll, MD, MPH,
Robert B. Thralls, Joseph A. Reid, PhD
Information Resources Management Office
Centers for Disease Control and Prevention, Atlanta, GA*

CDC WONDER, an information system developed at the Centers for Disease Control and Prevention (CDC), provides access to 26 text and numeric databases, and special facilities for surveillance, through an architecture developed for public health. We report extensions of the original architecture that allowed us to create a Web version (http://wonder.cdc.gov). *Keywords: Computers, Communications, Database, Informatics, Public Health.*

INTRODUCTION

CDC WONDER, an information system developed at the Centers for Disease Control and Prevention (CDC), provides access to 26 text and numeric databases through an architecture developed for public health (1). Users include: public health practitioners in Federal, state, and local government agencies who need rapid access to information in the form of pre-formatted tables of statistics and text; researchers who need 'raw" data for their own analyses; and surveillance and study coordinators who communicate data. In state and local health departments (where 50% of our audience works) one person may play several roles. A central goal for CDC WONDER is for a user to obtain information without needing to know the structure or location of the data.

The first version of CDC WONDER (released January, 1990) required dial-up access to the CDC mainframe (2). This made it somewhat cumbersome to use and download the results. The second version — CDC WONDER for the PC (released January, 1993) — had a DOS-client, easy downloading, and built-in graphics (1), and so was easier to use. It also provided access to E-mail and a bulletin board-like function. Importantly, the architecture was extended to allow access to data on a variety of platforms. The limitations of the PC system included a modem requirement for access (which was often the rate limiting step in response time), and the DOS client itself, which because of the many functions the system performed (database extractions, graphing, E-mail, bulletin boards), was hard to master.

Hence, we turned our attention to speeding the connection to CDC, and to improving the interface. Although we considered building a standalone Windows / X-Windows /MacIntosh client (1), the advent of the World Wide Web (WWW provided a short-cut to faster communications *and* to moving the interface to a multi-platform GUI. However, using the WWW without completely redoing the databases required the development of a method to move the complex functions performed by our DOS client to systems located at CDC. This involved important architectural extensions to the system, which we now report.

SYSTEM ARCHITECTURE

A. Background

The original architecture of CDC WONDER for the PC (1) was extended to allow access to the same data system via the WWW (a web browser that supports HTML forms and tables is required). No changes were made to the fundamental architecture of cooperative processing and pure message passing. Cooperative processing is to be distinguished from the traditional client-server model, wherein applications are divided between server and client, with both portions operating in tight integration. Rather, the components of WONDER reside on several platforms and are connected via simple messages which are passed by communications protocols that are ignorant of their contents. In fact, at all steps the communications protocols

* Product names are for identification only; their use does not constitute endorsement.

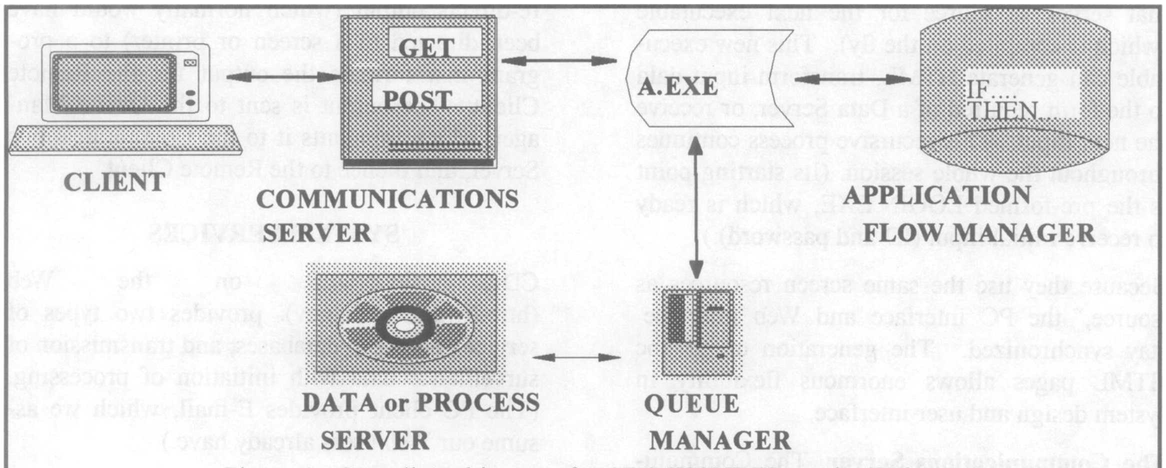

Figure 1. Overall Architecture for CDC WONDER on the Web;
A.EXE represents a sample executable (e.g., mortality request).

simply copy messages to readers. Only the reader can "open" a message. This architecture allows CDC WONDER to have data reside on many platforms, including an IBM mainframe under ESA for multi-gigabyte national databases, NT-based Structured Query Language (SQL) servers; and DOS systems.

This very architecture, because it is based on "loosely coupled" components, allowed it to be extended to allow Web access; *no changes* were made to the databases. In essence, we swapped-out the Client. (Incidental changes were also made to the back-end, to take advantage of the very fast Internet Protocol (IP) access to our mainframe.)

B. System Components

WONDER has five classes of components: the Remote Client; the Communications Servers; the Application Flow Manager (new); the Queue Managers; and Data Servers and Process Servers. (1) The new Application Flow Manager (AFM) is the key to the new Client, in that it replaces many of the intelligent functions previously handled by the DOS client, but which Web browsers currently cannot perform.

The Remote Client software resides on the user's machine; the other components are hardware and software agents on CDC's Wide Area Network (WAN). The Remote Client, Application Flow Manager, Communications Servers, and Queue Managers are the means by which WONDER allow users to access the Data and Process Servers (Figure 1). A typical Data Server is a numerical or text database; a typical Process Server is a surveillance system. In the original architecture, the components were linked by various communications protocols; in CDC WONDER on the Web, the client is always linked via IP. Discussion of the components will focus on changes and extensions that allow Web access to the (unchanged) Data and Process Servers.

The Remote Client and Application Flow Manager. CDC WONDER on the Web uses a standard Web browser as the Remote Client, and it can be run on any OS (Windows, MacIntosh, UNIX, etc.) This — plus standard full-screen graphics access to the WWW — is all a user needs to access the system.

WONDER generates HTML code *on the fly* to produce screens based on previous input or output. For example, to complete a request for mortality data for Fulton County, Georgia, one first specifies "Georgia"; the next screen displays a list of the counties in Georgia. Similarly, a request for *Morbidity and Mortality Weekly Report* articles that contain the words "Tuberculosis" AND "Children" generates a list of articles from a database that is updated every week; the user can then select specific articles to view.

The HTML pages are generated by AFM executables that are themselves created on the fly. The rules that guide their generation are in the AFM database (Figure 1). In essence, input (e.g., "Georgia") is sent to an AFM executable, which looks up a rule in its database (IF Georgia, THEN County_List=GA.Resource). This rule points to the Georgia portion of a pre-formed (physical) resource file (State.Resource) that serves as source for the next executable (which is compiled on the fly). This new executable can generate HTML, transform input data to the form required of a Data Server, or receive the next input. This recursive process continues throughout the whole session. (Its starting point is the pre-formed LOGIN.EXE, which is ready to receive initial input (ID and password).)

Because they use the same screen resources as “source,” the PC interface and Web interface stay synchronized. The generation of ad hoc HTML pages allows enormous flexibility in system design and user interface.

The Communications Server. The Communications Server is a modified version of the CERN HTTPD server (HyperText Transfer Protocol Daemon). It has two basic functions, which the *client* initiates: GET data *from* and POST *to* the AFM executables. Two enhancements were made to the CERN server: 1) the client can initiate execution of AFM executable (after passing data); and 2) functions related to identification and management of an individual user were enhanced, thereby allowing us to build a stand-alone database of users, and thus to grant specific rights (say, access to specialized data) to different users.

Queue Managers. Each Queue Manager handles the exchange of files between the Communications Server and those Data and Process Servers not accessible via File Transfer Protocol (FTP), including systems on our Novell WAN, which relies on IPX/SPX. Queue Managers queue, track, optionally reformat, and transmit files; and start Data or Process Servers. Those Data and Process servers that are directly accessible via FTP rely on their own queuing functions (e.g., the mainframe's job entry queue.)

Data and Process Servers. Data Servers are database applications in Natural/ADABAS and SAS on the IBM mainframe, in SQL databases under OS/2, and in FoxBase under DOS. Any kind of database application that can be accessed via IP or Novell IPX/SPX can be supported. Process Servers respond to input without necessarily performing any database functions (e.g., gateways; programs that send notification of the receipt data).

To access information, the Client creates a file that is used to drive a Data or Process Server as if keystrokes were being entered, but the Server re-directs output (which normally would have been directed to a screen or printer) to a program that formats the output for the Remote Client. The output is sent to the Queue Manager, which transmits it to the Communications Server, and thence to the Remote Client.

SYSTEM SERVICES

CDC WONDER on the Web (http://wonder.cdc.gov). provides two types of service: access to databases; and transmission of surveillance data with initiation of processing. (The PC client provides E-mail, which we assume our Web users already have.)

A. Access to Databases

There are 26 numerical and text databases (Table 1-2).

Table 1. Condition- and Risk Factor-Specific Databases, CDC WONDER [*]

- ♦ AIDS Surveillance Data (Disease Detail), 1983-1994
- ♦ AIDS Surveillance Data (Geographic Detail), 1983-1993
- ♦ Cancer Surveillance, Epidemiology & End Results, 1973-89[†]
- ♦ Fatal Accident Reporting System, 1991 [‡]
- ♦ Fluoridation Programs
- ♦ Natality and Maternal Characteristics, 1992
- ♦ National Hospital Discharge Survey, 1985 -90
- ♦ Occupational Mortality by Occupation Groups
- ♦ Occupational Mortality, 1990-94
- ♦ Population by Age, Race, Gender, County, 1970-92[¥]
- ♦ Sexually Transmitted Diseases Surveillance, 1989-1994
- ♦ Underlying Cause of Death, 1979-92

[*] All data from the Centers for Disease Control and Prevention, unless otherwise specified.
[†] National Institutes of Health
[‡] National Highway Traffic Safety Administration
[¥] Bureau of the Census

Data are accessed by ‘fill-in-the-blank” forms. Generally, summary information only is available, categorized by 2-4 keys; the user specifies which of these keys to invoke for reporting (e.g., show the data by age and year). Results are typically returned in 30-90 seconds. For large requests, the user may have results E-mailed at a later time; data can be formatted several ways, including ASCII, DIF, and comma delimited.

Table 2. General Reference Databases in CDC WONDER *

- Advisories for Overseas Travelers
- Calendar of Courses at CDC and Elsewhere
- CDC Prevention Guidelines
- CDC Publications (title, author, abstract)
- Chronic Disease Prevention Bibliography
- Cost Benefit/Cost Effectiveness Bibliography[3]
- Healthy People 2000 Objectives, Data
- Health Reform Activities
- Injury Resources
- International Classification of Diseases: Morbidity†
- International Classification of Diseases: Mortality ‡
- *Morbidity & Mortality Weekly Report* (full text), 1982-on
- Primary Care Models That Work
- Sexually Transmitted Disease Documents

* All data from Centers for Disease Control and Prevention, unless otherwise specified.

† In collaboration with the World Health Organization

‡ Health Resources Services Agency

B. Surveillance Data

Surveillance databases are defined by CDC staff and collaborators in state and local health departments. The Communications Server can perform a FTP file transfer from the user's machine (it must be triggered by the Remote Client). The data can be a pre-formed file on the workstation, or entered on-line. For WONDER to initiate the transfer, the user must specify an ID, password, *and* the IP address of the workstation with the data; this information must already have been entered in WONDER's security database. The ID and password can be set for one-time use. (We urge that the data be encrypted.)

Once the data arrive at CDC, WONDER can transfer the data to anywhere on the CDC WAN, and processes such as updating a database or generating a report. In this fashion, surveillance data can be made available for instant analysis by a restricted group, or to the public. Currently (March, 1996), security concerns have prevented CDC surveillance data from being transmitted over the Internet. However, there are a number of offices evaluating the use of CDC WONDER on the Web for surveillance.

SYSTEM UTILIZATION

CDC WONDER on the Web (released December, 1995) is currently (March, 1996) accessed approximately 10,000 times per week, which is already exceeding the usage of the PC system (exclusive of E-mail). The vast majority of inquiries are targeted at the CDC Prevention Guidelines, a collection of more than 400 key documents (Figure 2).[4] It is searchable by subject (86% of total usage) or by free text search (5% of usage). The other data, which are more technical, have received much less usage.

DISCUSSION

The fundamental architectural feature of CDC WONDER is the coupling of many independent components via cooperative processing and pure message-passing.

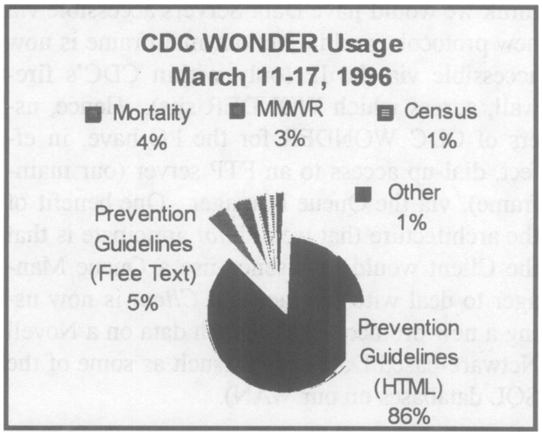

Figure 2. System Utilization.

This paradigm is a logical extension of other distributed computing environments (5), and has allowed us to "grow" the system in directions that we did not originally anticipate — in particular, adding an Internet client. The architecture has not been changed, but rather, extended.

From the user's perspective, CDC WONDER on the Web facilitates access to national information resources and the timely transmittal of data. For the application developer, the WONDER architecture provides great flexibility in the choice of hardware, operating and communications systems, and applications software; and so provides an economical migration path for developers to provide an Internet front-end to existing mainframe or minicomputer applications.

It is important to distinguish the WONDER architecture from "middle-ware" and "gateways", which are based on data *translations*. While they can be valuable interim strategies, they require continual maintenance, and so tend not to be robust. Moreover, experience tells that sooner or later, one of the 'languages" will incorporate unannounced or untranslatable changes. By contrast, WONDER is language and protocol independent; all the systems are loosely coupled; there is no native or foreign system; all systems are peer.

Has our previous prediction of the extensibility afforded by the WONDER architecture been born out? (1) We had noted that having a Queue Manager disconnected from the processes it serves ought to make it possible to add Data Servers residing on a wide variety of protocols. This has turned out to be true in two ways — one we anticipated, and one we did not. We did think we would have Data Servers accessible via new protocols, and in fact our mainframe is now accessible via the IP (only within CDC's firewall, across which WONDER sits). Hence, users of CDC WONDER for the PC have, in effect, dial-up access to an FTP server (our mainframe), via the Queue Manager. One benefit of the architecture that we did *not* anticipate is that the Client would in essence use a Queue Manager to deal with the fact that *Client* is now using a new protocol (IP) to reach data on a Novell Netware-based Data Server (such as some of the SQL databases on our WAN).

The most important benefit of this architecture is that we were able to add access via the Internet to 26 databases, without modifying them. Thus, there is only one set of databases and screen sources to maintain, so the two versions of WONDER (PC and Web) are always synchronized. Another important result is that our partners in state and local health departments (many of whom do not have access to the Web, or are just now obtaining it) will automatically have access to new and modified databases. The GUI interface, and better response time, are inducing health departments to obtain Web access.

We evaluated ways to speed-access to large dataset applications: the data transfer from the LAN-based Queue Manager to the mainframe-based Data Server; and the need to manually request responses. CDC WONDER on the Web has halved the mean response time as compared to the CDC WONDER for the PC, even though the overall load on the (shared) databases has increased several-fold, by eliminating the Queue Manager for data now accessible via IP (the mainframe); and by automating the downloading of results to the Client.

Future plans call for simplifying the process whereby other health agencies can add data to the system. This is vital for public health's viability in the era of health reform.[6,7] and involves developing ways to allow others to hook to WONDER, but retain the data locally. This work — which we call 'Distributed WONDER" — is underway.

REFERENCES

1. Friede A, Rosen DH, Reid JA. CDC WONDER/PC: cooperative processing for public health informatics. *J Am Med Informatics Assoc.* 1994; 1:303-312.
2. Friede A, Reid JA, Ory, HW. CDC WONDER: a comprehensive online public health information system of the Centers for Disease Control and Prevention. *Am J Public Health.* 1993;83(8):1289-1294.
3. Friede A, Taylor WR, Nadelman, L. Online access to A CBA/CEA bibliography via CDC WONDER. *Med Care.* 1993;31:JS12-JS17.
4. Friede A, O'Carroll PW, Nicola RM, Oberle MW, Teutch SM. *The CDC Prevention Guidelines.* Baltimore: Williams & Wilkins, 1996. (In press.)
5. Pine R, Presotto D, Thompson K, Trickey H. Designing Plan 9. *Dr. Dobb's Journal.* 1991; January:49-60.
6. Friede A, McDonald MC, Blum H. Public health informatics: how information-age technology can strengthen public health. *Ann Rev Public Health.* 1995; 16:239-252.
7. Baker EL, Friede A, Moulton AD, Ross DA. CDC's Information Network for Public Health Officials (INPHO): A framework for integrated public health information and practice. *J Public Health Management Practice.* 1995; 1:43-47.

ACKNOWLEDGMENTS

We thank the staff of the Public Health Information Systems Branch for developing CDC WONDER. We are indebted to colleagues who provided data, and associated report designs.